

Ruby - Feature #2542

URI lib should be updated to RFC 3986

01/01/2010 03:45 AM - marcandre (Marc-Andre Lafortune)

Status:	Closed	
Priority:	Normal	
Assignee:	naruse (Yui NARUSE)	
Target version:	2.6	
Description =begin RFC 2396 has been obsolete for nearly 5 years now. It was replaced by RFC 3986 which aims at clarifying aspects that were not previously clear. =end		
Related issues:		
Related to Ruby - Bug #4110: WEBrickError	Closed	12/02/2010
Related to Ruby - Bug #4673: URI::Generic registry is not properly set.	Rejected	
Related to Ruby - Bug #8352: URI squeezes a sequence of slashes in merging pa...	Closed	
Related to Ruby - Bug #10402: URI regression in 2.2.0-preview1 (bad URI(is no...	Closed	
Related to Ruby - Bug #9990: URI.parse and URI.encode use different RFCs	Closed	
Related to Ruby - Bug #21294: URI.extract is extracting invalid URIs with a m...	Open	

Associated revisions

Revision bb83f32dc3e0424d25fa4e55d8ff32b061320e41 - 06/22/2014 12:22 AM - naruse (Yui NARUSE)

support RFC3986 [Feature #2542]

- lib/uri/common.rb (URI::REGEXP): move to lib/uri/rfc2396_parser.rb.
- lib/uri/common.rb (URI::Parser): ditto.
- lib/uri/common.rb (URI.split): use RFC3986_Parser.
- lib/uri/common.rb (URI.parse): ditto.
- lib/uri/common.rb (URI.join): ditto.
- lib/uri/common.rb (URI.extract): deprecated.
- lib/uri/common.rb (URI.regexp): ditto.
- lib/uri/rfc2396_parser.rb: added.
- lib/uri/rfc3986_parser.rb: added.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@46491 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision bb83f32d - 06/22/2014 12:22 AM - naruse (Yui NARUSE)

support RFC3986 [Feature #2542]

- lib/uri/common.rb (URI::REGEXP): move to lib/uri/rfc2396_parser.rb.
- lib/uri/common.rb (URI::Parser): ditto.
- lib/uri/common.rb (URI.split): use RFC3986_Parser.
- lib/uri/common.rb (URI.parse): ditto.
- lib/uri/common.rb (URI.join): ditto.
- lib/uri/common.rb (URI.extract): deprecated.
- lib/uri/common.rb (URI.regexp): ditto.

- lib/uri/rfc2396_parser.rb: added.
- lib/uri/rfc3986_parser.rb: added.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@46491 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 7af73882028f65c96015a1c8801b40565f9c0119 - 07/03/2014 06:00 PM - naruse (Yui NARUSE)

- lib/uri/generic.rb (URI::Generic#query=): remove validation, just escape. [Feature #2542]
- lib/uri/generic.rb (URI::Generic#fragment=): ditto.
- lib/uri/generic.rb (URI::Generic#check_query): removed.
- lib/uri/generic.rb (URI::Generic#set_query): ditto.
- lib/uri/generic.rb (URI::Generic#check_fragment): ditto.
- lib/uri/generic.rb (URI::Generic#set_fragment): ditto.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@46680 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 7af73882 - 07/03/2014 06:00 PM - naruse (Yui NARUSE)

- lib/uri/generic.rb (URI::Generic#query=): remove validation, just escape. [Feature #2542]
- lib/uri/generic.rb (URI::Generic#fragment=): ditto.
- lib/uri/generic.rb (URI::Generic#check_query): removed.
- lib/uri/generic.rb (URI::Generic#set_query): ditto.
- lib/uri/generic.rb (URI::Generic#check_fragment): ditto.
- lib/uri/generic.rb (URI::Generic#set_fragment): ditto.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@46680 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

History

#1 - 01/01/2010 03:46 AM - marcandre (Marc-Andre Lafortune)

- Subject changed from *URI lib should be updated to RFC 3986* to *URI lib should be updated to RFC 3986*

#2 - 01/02/2010 12:58 AM - naruse (Yui NARUSE)

FYI, RFC 3986 will be obsoleted.

<http://tools.ietf.org/html/draft-duerst-iri-bis-07>

#3 - 01/05/2010 03:11 PM - duerst (Martin Dürst)

No, RFC 3986 (URI) will NOT be updated. RFC 3987 (IRI), in due time, will be updated. See also

<http://www.ietf.org/ibin/c5i?mid=6&rid=49&qid=0&k1=934&k2=7294&tid=1262671757>.

Regards, Martin.

Martin J. Dürst, Professor, Aoyama Gakuin University

<http://www.sw.it.aoyama.ac.jp> <mailto:duerst@it.aoyama.ac.jp>

#4 - 01/06/2010 10:10 AM - naruse (Yui NARUSE)

Ah, Martin is right, it's RFC 3987 and RFC 3986 will be still alive.

Anyway Bob Aman introduces an alternative library named Addressable.
<http://addressable.rubyforge.org/>
It looks good but some incompatibilities for current URI library.
I'll suggest to bundle Addressable and obsolete current URI lib,
but I have to plan its migration path.

#5 - 03/18/2010 02:12 AM - mame (Yusuke Endoh)

- Target version changed from 1.9.2 to 2.0.0

Hi,

I'll suggest to bundle Addressable and obsolete current URI lib,
but I have to plan its migration path.

This ticket seems to need much work.
I guess we can't make the deadline of spec freezing.
So I change the target to 1.9.x.
If you want 1.9.2 to include the feature, please discuss right now.

Yusuke Endoh mame@tsg.ne.jp

#6 - 03/25/2010 11:39 AM - marcandre (Marc-Andre Lafortune)

- Target version changed from 2.0.0 to 1.9.2

I feel the spec for 1.9.2 has been quite clear for 5 years ... follow RFC 3986!

Integrating some of the features of the addressable gem can be discussed later.

Do we have to wait for Akira Yamada, the official maintainer of this library?

#7 - 03/25/2010 11:56 AM - naruse (Yui NARUSE)

I object to target 1.9.2.
Following RFC 3986 makes some incompatibilities.
It shouldn't be done without consideration.

#8 - 03/25/2010 01:25 PM - duerst (Martin Dürst)

Hello Yui,

Is there a list of incompatibilities, or can you make one?

Regards, Martin.

#-# Martin J. Dürst, Professor, Aoyama Gakuin University
#-# <http://www.sw.it.aoyama.ac.jp> <mailto:duerst@it.aoyama.ac.jp>

#9 - 03/25/2010 04:35 PM - naruse (Yui NARUSE)

2010/3/25 "Martin J. Dürst" duerst@it.aoyama.ac.jp:

Is there a list of incompatibilities, or can you make one?

Some structures of the syntax is changed in RFC 3986.
This breaks URI::REGEXP::PATTERN::TOPLABEL and some constants.

NARUSE, Yui
naruse@airemix.jp

#10 - 04/04/2010 01:31 AM - znz (Kazuhiro NISHIYAMA)

- Status changed from Open to Assigned

- Target version changed from 1.9.2 to 2.0.0

#11 - 05/13/2010 08:01 AM - hedgehog (Hedge Hog)

Rather than reinvent anything. Consider employing an FFI interface to uriparser:

<http://uriparser.sourceforge.net/>

not sure if there is a port for windows or if an equivalent windows lib is available?

#12 - 05/13/2010 09:05 AM - luislavena (Luis Lavena)

On Wed, May 12, 2010 at 8:01 PM, Hedge Hog redmine@ruby-lang.org wrote:

Issue [#2542](#) has been updated by Hedge Hog.

Rather than reinvent anything. Consider employing an FFI interface to uriparser:

<http://uriparser.sourceforge.net/>

not sure if there is a port for windows or if an equivalent windows lib is available?

we will not only depend on uriparser C library but also will require libcpptest to be able to configure and compile uriparser.

Luis Lavena
AREA 17

Perfection in design is achieved not when there is nothing more to add,
but rather when there is nothing more to take away.
Antoine de Saint-Exupéry

#13 - 09/23/2010 11:31 PM - philo (Philippe Lucas)

It depends of libcpptest only for test so you can build the package with '--disable-test'.

#14 - 12/03/2010 03:07 AM - naruse (Yui NARUSE)

- Assignee changed from akira (akira yamada) to naruse (Yui NARUSE)

I come to think uri lib should move to RFC 3986 even if it breaks some compatibility.
But I don't want that new implementation/spec will be also a white box like now.
So I think:

- keep current URI::REGEXP, URI::Parser and so on.
- at least URI.parse doesn't use URI::Parser but use new implementation.

How about this?

#15 - 07/16/2011 10:40 PM - naruse (Yui NARUSE)

- Target version changed from 2.0.0 to 1.9.4

Just for information or memo.

<http://tools.ietf.org/html/draft-ietf-iri-3987bis-05>

<http://tools.ietf.org/html/draft-weber-iri-guidelines-00>

#16 - 11/20/2012 05:48 PM - nikosd (Nikos Dimitrakopoulos)

Are there any plans for actually fixing this? Not sure I can help, and no troll appetite - just asking :)

#17 - 11/24/2012 11:22 AM - mame (Yusuke Endoh)

- Target version changed from 1.9.4 to 2.6

Naruse-san, could you please answer to Nikos?

I'm setting to next minor, but if you are willing to do anything to 2.0.0, and if the impact is so small, I may accept.

--

Yusuke Endoh mame@tsg.ne.jp

#18 - 05/29/2013 10:34 AM - naruse (Yui NARUSE)

Just an experimental implementation:

<http://github.com/nurse/url>

#19 - 06/22/2014 12:22 AM - naruse (Yui NARUSE)

- Status changed from Assigned to Closed
- % Done changed from 0 to 100

Applied in changeset r46491.

support RFC3986 [Feature [#2542](#)]

- lib/uri/common.rb (URI::REGEXP): move to lib/uri/rfc2396_parser.rb.
- lib/uri/common.rb (URI::Parser): ditto.
- lib/uri/common.rb (URI.split): use RFC3986_Parser.
- lib/uri/common.rb (URI.parse): ditto.
- lib/uri/common.rb (URI.join): ditto.
- lib/uri/common.rb (URI.extract): deprecated.
- lib/uri/common.rb (URI.regexp): ditto.
- lib/uri/rfc2396_parser.rb: added.
- lib/uri/rfc3986_parser.rb: added.

#20 - 06/30/2014 09:55 PM - tenderlovmaking (Aaron Patterson)

r46491 broke this script:

```
require 'uri'

thing = URI.parse 'http://example.com'
thing.query = 'location[]=1&location[]=2&age_group[]=2'
```

Before r46491 it would set the query, after r46491, it raises an exception.

Is this a bug in the new implementation? Or should I be doing something different? (I pulled this from the Rails tests, so I'm not 100% sure what it is actually for)

#21 - 06/30/2014 10:24 PM - bitsweat (Jeremy Daer)

In RFC 3986, square brackets are no longer allowed in the query part.

Source of the unescaped brackets, in this case: <https://github.com/brynary/rack-test/blob/master/lib/rack/test/utls.rb>

This may become a common issue since plenty of code uses URI.parse and expects its more permissive RFC 2396 parsing.

#22 - 07/01/2014 05:59 AM - zzak (zzak _)

I think [#9990](#) is related /cc [@naruse \(Yui NARUSE\)](#) @JK @tenderlove

#23 - 07/01/2014 09:14 AM - naruse (Yui NARUSE)

I'm considering to change the error policy of URI library, for example:
BEFORE: raise error if invalid characters exist
AFTER: percent-escape them

#24 - 07/01/2014 01:10 PM - lengarvey (Leonard Garvey)

I've implemented something similar to that policy in the following gist: <https://gist.github.com/lengarvey/31983eac6664351ed16d>

It's a very basic naive implementation but I believe it roughly does what we need URI.parse to do.

#25 - 10/20/2014 12:37 AM - nobu (Nobuyoshi Nakada)

- Related to Bug #10402: URI regression in 2.2.0-preview1 (bad URI(is not URI?): URI::InvalidURIError) added

#26 - 10/20/2014 12:38 AM - nobu (Nobuyoshi Nakada)

- Related to Bug #9990: *URI.parse* and *URI.encode* use different RFCs added

#27 - 05/07/2025 07:56 AM - mame (Yusuke Endoh)

- Related to Bug #21294: *URI.extract* is extracting invalid URLs with a mishmash of IPv6 notation with IPv4 address added